

# Approximability of the robust representatives selection problem

Adam Kasperski

*Institute of Industrial  
Engineering and Management,  
Wrocław University of Technology,  
Wybrzeże Wyspiańskiego 27,  
50-370 Wrocław, Poland,  
adam.kasperski@pwr.edu.pl*

Adam Kurpisz, Paweł Zieliński\*

*Institute of Mathematics  
and Computer Science  
Wrocław University of Technology,  
Wybrzeże Wyspiańskiego 27,  
50-370 Wrocław, Poland  
{adam.kurpisz,pawel.zielinski}@pwr.edu.pl*

## Abstract

In this paper new complexity and approximation results on the robust versions of the representatives selection problem, under the scenario uncertainty representation, are provided, which extend the results obtained in the recent papers by Dolgui and Kovalev (2012), and Deineko and Woeginger (2013). Namely, it is shown that if the number of scenarios is a part of input, then the min-max (regret) representatives selection problem is not approximable within a ratio of  $O(\log^{1-\epsilon} K)$  for any  $\epsilon > 0$ , where  $K$  is the number of scenarios, unless the problems in NP have quasi-polynomial time algorithms. An approximation algorithm with an approximation ratio of  $O(\log K / \log \log K)$  for the min-max version of the problem is also provided.

**Keywords:** robust optimization, selection problem, uncertainty, computational complexity

## 1 Preliminaries

In [3, 6] the min-max and min-max regret versions of the following *representatives selection* problem (RS for short) have been recently discussed. We are given a set  $T$  of  $n$  tools, numbered from 1 to  $n$ . This set is partitioned into  $p$  disjoint sets  $T_1, \dots, T_p$ , where  $|T_i| = r_i$  and  $n = \sum_{i \in [p]} r_i$  (we use  $[p]$  to denote the set  $\{1, \dots, p\}$ ). We wish to choose a subset  $X \subseteq T$  of the tools that contains exactly one tool from each set  $T_i$ , i.e.  $|X \cap T_i| = 1$  for each  $i \in [p]$ . In the deterministic case, each tool  $j \in T$  has a nonnegative cost  $c_j$  and we seek a solution  $X$  whose total cost  $F(X) = \sum_{j \in X} c_j$  is minimal. This problem can be solved by a trivial algorithm, which chooses a tool of the smallest cost from each  $T_i$ . An important problem characteristic is the maximal number of elements in  $T_i$ , i.e.  $r_{\max} = \max_{i \in [p]} r_i$ . We can assume that  $r_i \geq 2$  for all  $i \in [p]$ , since a subset with the cardinality of 1 can be ignored. It is easy to see that RS is a special case of the SHORTEST PATH problem, and this fact is depicted in Figure 1. Each solid arc in the graph  $G$  shown in Figure 1 corresponds to a tool in  $T_i$  and each dummy (dashed) arc has zero cost. There is one-to-one correspondence between the  $s$ - $t$  paths in  $G$  and the solutions to RS. It is not difficult to transform the network  $G$  and show that RS is also a special case of other basic network problems such as: MINIMUM SPANNING TREE, MINIMUM ASSIGNMENT, or MINIMUM CUT (see, e.g. [7]).

---

\*Corresponding author

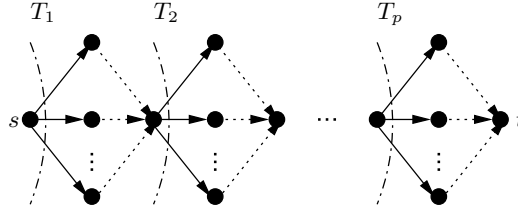


Figure 1: The representative selection problem as the  $s$ - $t$  shortest path problem.

Following [6], let us now define the robust versions of RS. Let  $\Gamma = \{S_1, \dots, S_K\}$  be a *scenario set*, where each *scenario* is a vector  $S_k = (c_{k1}, \dots, c_{kn})$  of nonnegative integral tool costs. Let  $F(X, S_k) = \sum_{j \in X} c_{kj}$  be the cost of solution  $X \in \Phi$  under scenario  $S_k$  and let  $F^*(S_k)$  be the cost of an optimal solution under  $S_k$ . In order to choose a solution, two robust criteria, called the *min-max* and the *min-max regret*, can be adopted (see [8] for a motivation of both robust criteria). In the MIN-MAX RS problem, we seek a solution which minimizes the largest cost over all scenarios, that is

$$OPT_1 = \min_{X \in \Phi} cost_1(X) = \min_{X \in \Phi} \max_{k \in [K]} F(X, S_k). \quad (1)$$

In the MIN-MAX REGRET RS problem, we wish to find a solution which minimizes the *maximal regret*, that is

$$OPT_2 = \min_{X \in \Phi} cost_2(X) = \min_{X \in \Phi} \max_{k \in [K]} (F(X, S_k) - F^*(S_k)).$$

We now describe the known results on MIN-MAX (REGRET) RS. Notice first that, all positive results for MIN-MAX (REGRET) SHORTEST PATH (see, e.g. [1]) remain valid for MIN-MAX (REGRET) RS, which is due to the fact that there is one-to-one correspondence between  $s$ - $t$  paths in a layered digraph and solutions to RS (see Figure 1). Hence the latter problem admits a simple  $K$ -approximation algorithm which outputs an optimal solution for the aggregated tool costs  $\hat{c}_j = \max_{k \in [K]} c_{kj}$ ,  $j \in [n]$ . Furthermore, if  $K$  is constant, then MIN-MAX (REGRET) RS can be solved in pseudopolynomial time and admits an FPTAS. It has been shown in [6] that MIN-MAX (REGRET) RS is NP-hard even when  $K = 2$  and  $r_i = 2$  for all  $i \in [p]$ , and becomes strongly NP-hard when the number of scenarios is a part of the input. This result has been recently extended in [3], where it has been shown that when  $r_i = 2$  for all  $i \in [p]$ , then MIN-MAX (REGRET) RS is as hard to approximate as the vertex cover problem, which is conjectured to be hard to approximate within  $2 - \epsilon$  for any  $\epsilon > 0$ . A similar inapproximability result, by assuming  $P \neq NP$ , follows immediately from the reduction given in [7] for the MIN-MAX (REGRET) SHORTEST PATH problem. However, in the instances from [7] we have  $r_i = 3$  for all  $i \in [p]$ , so the result obtained in [3] is stronger.

**Our results** In this paper we investigate the case when both  $K$  and  $r_{\max}$  are parts of the input. We show that MIN-MAX RS admits an  $r_{\max}$ -approximation algorithm whose idea is to round up solutions computed by solving a linear relaxation. In particular, it admits a 2-approximation algorithm when  $r_{\max} = 2$  which is the best possible according to the negative results given in [3]. We also show, and these are the main results of the paper, that when additionally  $r_{\max}$  is a part of the input, then MIN-MAX (REGRET) RS is not

approximable within  $O(\log^{1-\epsilon} K)$  for any  $\epsilon > 0$  unless  $\text{NP} \subseteq \text{DTIME}(n^{\text{poly}(\log n)})$  and provide an approximation algorithm with a performance ratio of  $O(\log K / \log \log K)$  for MIN-MAX RS, which is close to the above lower bound.

## 2 Hardness result

In order to establish the hardness result, we will use the following variant of the LABEL COVER problem (see e.g., [2, 9]):

**LABEL COVER:** We are given a regular bipartite graph  $G = (V \cup W, E)$ ,  $E \subseteq V \times W$ , a set of labels  $[N]$ , and for each edge  $(v, w) \in E$  a map (partial)  $\sigma_{v,w} : [N] \rightarrow [N]$ . A *labeling* of  $G$  is an assignment of a subset of labels to each of the vertices of  $G$ , i.e. a function  $l : V \cup W \rightarrow 2^{[N]}$ . We say that a labeling *satisfies* an edge  $(v, w) \in E$  if there exist  $a \in l(v)$  and  $b \in l(w)$  such that  $\sigma_{v,w}(a) = b$ . A *total labeling* is a labeling that satisfies all edges. We seek a total labeling whose value defined as  $\max_{x \in V \cup W} |l(x)|$  is minimal. This minimal value is denoted by  $\text{val}(\mathcal{L})$ , where  $\mathcal{L}$  is the input instance.

**Theorem 1** ([9]). *There exists a constant  $\gamma > 0$  such that for any language  $L \in \text{NP}$ , any input  $\mathbf{w}$  and any  $N > 0$ , one can construct a LABEL COVER instance  $\mathcal{L}$  with the following properties in time polynomial in the instance's size:*

- *the number of vertices in  $\mathcal{L}$  is  $|\mathbf{w}|^{O(\log N)}$ ,*
- *if  $\mathbf{w} \in L$ , then  $\text{val}(\mathcal{L}) = 1$ ,*
- *if  $\mathbf{w} \notin L$ , then  $\text{val}(\mathcal{L}) > N^\gamma$ .*

We now prove the following result:

**Theorem 2.** *There exists a constant  $\gamma > 0$  such that for any language  $L \in \text{NP}$ , any input  $\mathbf{w}$ , and any  $N > 0$ , one can construct an instance of MIN-MAX RS with the following properties:*

- *if  $\mathbf{w} \in L$ , then  $\text{OPT}_1 \leq 1$ ,*
- *if  $\mathbf{w} \notin L$ , then  $\text{OPT}_1 \geq \lfloor N^\gamma \rfloor := g$ ,*
- *the number of tools is at most  $|\mathbf{w}|^{O(\log N)} N$  and the number of scenarios is at most  $|\mathbf{w}|^{O(g \log N)} N^g$ .*

*Proof.* Let  $L$  be a language in NP and let  $\mathcal{L} = (G = (V \cup W, E), N, \sigma)$  be the instance of LABEL COVER constructed for  $L$  (see Theorem 1). For each edge  $(v, w) \in E$  we create a subset of tools  $T_{v,w}$ , which contains at most  $N$  tools labeled as  $(i, \sigma_{v,w}(i))$ ,  $i \in [N]$ . A sample reduction is shown in Figure 2.

A pair of tools  $(i_1, j_1) \in T_{v_1, w_1}$  and  $(i_2, j_2) \in T_{v_2, w_2}$  is *label distinct* if  $i_1 = i_2$  implies  $v_1 \neq v_2$  and  $j_1 = j_2$  implies  $w_1 \neq w_2$ . In other words, this pair cannot assign the same label twice to any vertex. For example, the pair  $(3, 1) \in T_{v_1, w_1}$  and  $(1, 1) \in T_{v_1, w_2}$  is label distinct, whereas the pair  $(1, 2) \in T_{v_1, w_1}$  and  $(1, 1) \in T_{v_1, w_2}$  is not, as it assigns label 1 twice to  $v_1$ . We are now ready to form scenario set  $\Gamma$ . Let us fix a vertex  $v \in V$ . For each  $g$ -tuple  $(v, w^1), \dots, (v, w^g)$  of pairwise distinct edges incident to  $v$  and for each  $g$ -tuple of pairwise

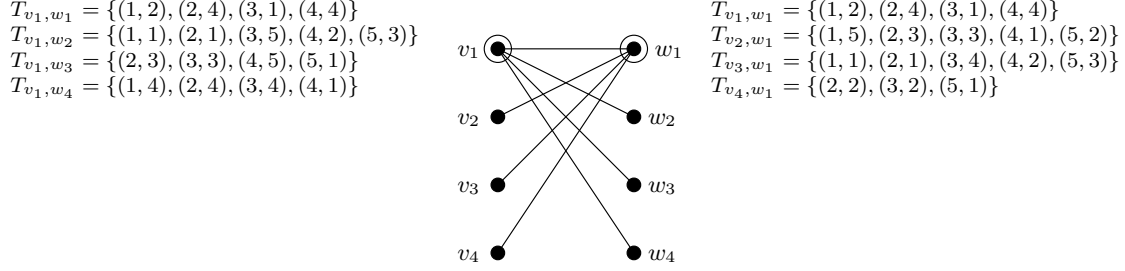


Figure 2: A sample reduction for  $K_{4,4}$  graph and  $N = 5$ . Only the edges incident to  $v_1$  and  $w_1$  and the subsets of tools corresponding to them are shown.

label distinct tools from  $T_{v, w^1} \times T_{v, w^2} \times \dots \times T_{v, w^g}$ , we form a scenario under which these tools have costs equal to 1 and all the remaining tools have costs equal to 0. We repeat this procedure for each vertex  $v \in V$ . A sample scenario for vertex  $v_1$  in Figure 2 and  $g = 3$  assigns the cost equal to 1 to tools  $(1, 2) \in T_{v_1, w_1}$ ,  $(3, 5) \in T_{v_1, w_2}$  and  $(4, 5) \in T_{v_1, w_3}$  as these tools forms a 3-tuple of pairwise label distinct tools. Let us fix a vertex  $w \in W$ . Now, for each  $g$ -tuple  $(v^1, w), \dots, (v^g, w)$  of pairwise distinct edges incident to  $w$  and for each  $g$ -tuple of pairwise label distinct tools from  $T_{v^1, w} \times T_{v^2, w} \times \dots \times T_{v^g, w}$ , we form a scenario under which these tools have costs equal to 1 and all the remaining tools have costs equal to 0. We repeat this procedure for each vertex  $w \in W$ . A sample scenario for vertex  $w_1$  in Figure 2 and  $g = 3$  assigns the cost equal to 1 to tools  $(1, 2) \in T_{v_1, w_1}$ ,  $(1, 5) \in T_{v_2, w_1}$  and  $(3, 4) \in T_{v_3, w_1}$  as these tools forms a 3-tuple of pairwise label distinct tools. To ensure that  $\Gamma \neq \emptyset$ , we create one additional scenario under which each tool has a cost equal to 0.

Suppose that  $\mathbf{w} \in L$ . Then  $\text{val}(\mathcal{L}) = 1$  and there exists a total labeling  $l$  which assigns one label  $l(x)$  to each vertex  $x \in V \cup W$ . Let us choose tool  $(l(v), l(w)) \in T_{v, w}$  for each  $(u, v) \in E$ . We thus get a feasible selection of the tools  $X$ . It is easy to see that  $F(X, S) \leq 1$  under each scenario  $S \in \Gamma$  and, consequently  $\text{OPT}_1 \leq 1$ . Assume that  $\mathbf{w} \notin L$ . Then  $\text{val}(\mathcal{L}) > N^\gamma$  which implies  $\text{val}(\mathcal{L}) > \lfloor N^\gamma \rfloor = g$ . Assume, by contradiction, that  $\text{OPT}_1 < g$ , so there is a feasible selection  $X$  such that  $F(X, S) < g$  under each scenario  $S \in \Gamma$ . Observe that  $X$  corresponds to the total labeling  $l$  which assigns labels  $i$  to  $v_i$  and  $j$  to  $w_j$  when the tool  $(i, j)$  is selected from  $T_{v_i, w_j}$ . From the construction of  $\Gamma$  it follows that  $l$  assigns less than  $g$  distinct labels to each vertex  $x \in V \cup W$ , since otherwise  $F(X, S) = g$  for some scenario  $S \in \Gamma$ . We thus have  $\text{val}(\mathcal{L}) < g$ , a contradiction.

Let us now estimate the size of the constructed instance of MIN-MAX RS with respect to  $|\mathbf{w}|$ . The number of tools is at most  $|E|N$  and the number of scenarios is bounded by  $|V||W|^g N^g + |W||V|^g N^g + 1$ . According to Theorem 1, the number of vertices (and also edges) in  $G$  is  $|\mathbf{w}|^{O(\log N)}$ . Hence the number of tools is at most  $|\mathbf{w}|^{O(\log N)} N$  and the number of scenarios is at most  $|\mathbf{w}|^{O(g \log N)} N^g$ .  $\square$

**Theorem 3.** MIN-MAX RS is not approximable within  $O(\log^{1-\epsilon} K)$ , for any  $\epsilon > 0$ , unless  $\text{NP} \subseteq \text{DTIME}(n^{\text{poly}(\log n)})$ .

*Proof.* Let  $\gamma$  be the constant from Theorem 2. Consider a language  $L \in \text{NP}$  and an input  $\mathbf{w}$ . Fix any constant  $\beta > 0$  and choose  $N = \lceil \log^{\beta/\gamma} |\mathbf{w}| \rceil$ . According to Theorem 2, we can construct an instance of MIN-MAX RS in which the number of scenarios  $K$  is asymptotically bounded by  $|\mathbf{w}|^{\alpha N^\gamma \log N} N^{N^\gamma}$  for some constant  $\alpha > 0$ ,  $\text{OPT}_1 \leq 1$  if  $\mathbf{w} \in L$  and  $\text{OPT}_1 \geq$

$\lfloor \log^\beta |\mathbf{w}| \rfloor$  if  $\mathbf{w} \notin L$ . We obtain  $\log K \leq \alpha N^\gamma \log N \log |\mathbf{w}| + N^\gamma \log N \leq \alpha' \log^{\beta+2} |\mathbf{w}|$  for some constant  $\alpha' > 0$  and sufficiently large  $|\mathbf{w}|$ . Therefore,  $\log |\mathbf{w}| \geq (1/\alpha') \log^{1/(\beta+2)} K$  and the gap is at least  $\lfloor \log^\beta |\mathbf{w}| \rfloor \geq \lfloor 1/\alpha' \log^{\beta/(\beta+2)} K \rfloor$ . Since the constant  $\beta > 0$  can be arbitrarily large, we get that the gap is  $O(\log^{1-\epsilon} K)$  for any  $\epsilon = 2/(\beta+2) > 0$ . Furthermore, the instance of MIN-MAX RS can be computed in  $O(|\mathbf{w}|^{\text{poly}(\log |\mathbf{w}|)})$  time, which completes the proof.  $\square$

**Corollary 1.** MIN-MAX REGRET RS is not approximable within  $O(\log^{1-\epsilon} K)$ , for any  $\epsilon > 0$ , unless  $NP \subseteq DTIME(n^{\text{poly}(\log n)})$ .

*Proof.* A reduction is almost the same as the one from Theorem 2. We only add to each  $T_{u,v}$  a dummy tool with 0 cost under each scenario  $S$  and one additional scenario  $S'$  under which all the dummy tools have a large cost (say  $g+1$ ) and the original tools have costs equal to 0. Thus, no dummy tool can be a part of an optimal solution and  $OPT_2 = OPT_1$ .  $\square$

### 3 Approximation algorithms

In this section, we provide some LP-based approximation algorithms for MIN-MAX RS. Let us fix parameter  $L > 0$  and let  $T(L) \subseteq T$  be the set of all the tools  $j \in T$  for which  $c_{kj} \leq L$  for all scenarios  $k \in [K]$ . Clearly,  $T(L) = \bigcup_{i \in [p]} T_i(L)$ , where  $T_i(L) = \{j \in T_i : \forall k \in [K] c_{kj} \leq L\}$  is a modified subset of  $T_i$ . Consider the following linear program:

$$\mathcal{LP}(L) : \sum_{j \in T(L)} c_{kj} x_j \leq L, \quad k \in [K], \quad (2)$$

$$\sum_{j \in T_i(L)} x_j = 1, \quad i \in [p], \quad (3)$$

$$x_j \geq 0, \quad j \in T(L), \quad (4)$$

$$x_j = 0, \quad j \notin T(L). \quad (5)$$

Minimizing  $L$  subject to (2)-(5) we obtain an LP relaxation of MIN-MAX RS. Let  $L^*$  denote the smallest value of the parameter  $L$  for which  $\mathcal{LP}(L)$  is feasible. Clearly,  $L^*$  is a lower bound on  $OPT_1$  and can be determined in polynomial time by using binary search. If  $\mathbf{x}^*$  is a feasible, fractional solution to  $\mathcal{LP}(L^*)$ , then constraints (3) imply  $x_j \geq 1/r_{\max}$  for at least one tool  $j \in T_i(L^*)$  for each  $i \in [p]$ . By choosing such a tool from  $T_i(L^*)$ , we get a solution  $X$  with  $\text{cost}_1(X)$  at most  $r_{\max} \cdot L^* \leq r_{\max} \cdot OPT_1$ . This leads to the following observation:

**Observation 1.** MIN-MAX RS admits an  $r_{\max}$ -approximation algorithm.

It is worth pointing out that the above algorithm is the best possible when  $r_{\max} = 2$ , according to the negative results provided in [3]. The following observation describes the integrality gap of the LP relaxation:

**Observation 2.** The LP relaxation has an integrality gap of at least  $\Omega(\log K / \log \log K)$ .

*Proof.* Let  $p > 0$  be an arbitrary integer. Fix  $n = p^2$ . Let  $|T_i| = p$  for every  $i \in [p]$ . We form a scenario set  $\Gamma$  as follows. For each  $p$ -tuple  $(e_1, \dots, e_p) \in [p]^p$ , where the component  $e_i \in [p]$  corresponds to the  $e_i$ -th element in subset  $T_i$ , we form scenario under which the tools indicated by  $(e_1, \dots, e_p)$  have costs equal to 1 and all the remaining tools have costs equal to 0. Thus,

$|\Gamma| = K = p^p$ . Consider a fractional solution  $x_j = 1/p$ , for all  $j \in T$ . This solution is feasible to  $\mathcal{LP}(L^*)$ , where  $L^* = 1$ . It is easy to notice that each integer solution for the instance constructed has the maximum cost over all scenarios equal to  $p$ . Therefore, we have the integrality gap of  $p$ . Since  $\log K = p \log p$ ,  $\log K / \log \log K = p \log p / (\log p + \log \log p) = \Theta(p)$  and the integrality gap of the LP relaxation is at least  $\Omega(\log K / \log \log K)$ .  $\square$

We now convert a feasible, fractional solution  $\mathbf{x}^*$  to  $\mathcal{LP}(L^*)$  into an integer solution  $\mathbf{z} \in \{0, 1\}^n$ , which will represent a feasible tool selection. To do this a randomized rounding technique proposed in [11, 12] can be applied. We first remove all the tools whose cost under some scenario is greater than  $L^*$ , i.e.  $j \notin T(L^*)$ . We lose nothing by assuming, from now on, that the remaining tools  $j$ ,  $j \in T(L^*)$ , are numbered from 1 to  $n$ ,  $|T(L^*)| = n$ , and thus  $|T_i(L^*)| = r_i$  and  $n = \sum_{j \in [p]} r_i$ . Furthermore, we make the assumption that  $L^* = 1$  and all the tool costs are such that  $c_{kj} \in [0, 1]$ ,  $k \in [K]$ ,  $j \in [n]$ . One can easily meet this assumption by dividing all the tool costs  $c_{kj}$  by  $L^*$  for all  $k \in [K]$  and  $j \in [n]$ . It turns out that the above instance of the MIN-MAX RS problem with the scaled costs is an instance of the *minimax integer program* discussed in [12]. Therefore, we now construct  $\mathbf{z}$  as follows (see [11, 12]): independently, for each  $i \in [p]$ , pick one item  $\bar{j}$  from  $T_i$  with probability  $x_{\bar{j}}^*$ , set  $z_{\bar{j}} = 1$  and  $z_j = 0$  for the remaining items  $j \in T_i$ . It is clear that  $\mathbf{z}$  represents a feasible tool selection. Let  $\mathbf{C}$  denote the matrix corresponding to the scenario constraints (2), i.e.  $\mathbf{C} = (c_{kj}) \in [0, 1]^{K \times n}$  and  $(\mathbf{C}\mathbf{z})_k = \sum_{j \in [n]} c_{kj} z_j$  stands for the l.h.s value of the scenario constraint  $k$  for solution  $\mathbf{z}$ . It holds  $\mu_k = \mathbb{E}[(\mathbf{C}\mathbf{z})_k] = (\mathbf{C}\mathbf{x}^*)_k \leq L^* = 1$  for all  $k \in [K]$ , due to the linearity of expectation. We can make the assumption that  $\mu_k = 1$  for all  $k \in [K]$ , which can be satisfied by adding slack variables to the constraints (2). Now, an analysis similar to that in [12, the proof of Theorem 6.1] shows that  $(\mathbf{C}\mathbf{z})_k$  is at most  $O(\log K / \log \log K)$  for each  $k \in [K]$  with high probability. We thus get a randomized  $O(\log K / \log \log K)$  approximation algorithm for the MIN-MAX RS problem.

In order to derandomize the above algorithm, one can use the *method of pessimistic estimators* [10]. However, this method assumes that one can perform the computations with real numbers (in particular exponentials) with high precision in constant time. Thus, it fails in the RAM model, when  $\mathbf{C} \in [0, 1]^{K \times n}$  (see [10], for comments). However, it works in the RAM model when  $\mathbf{C} \in \{0, 1\}^{K \times n}$ . Note that this particular case, i.e. MIN-MAX RS with a binary matrix, is equivalent to the *vector selection* problem studied in [10]. We can thus formulate the analogue of results given in [10]. The following formulation, which provides bounds on rounding errors, will be very useful in a derandomization of the randomized algorithm in the RAM model.

**Theorem 4.** *Let  $\mathbf{C} \in \{0, 1\}^{K \times n}$  and let  $\mathbf{x} \in ([0, 1] \cap \mathbb{Q})^n$  be any solution that satisfies constraints (3). Then in the RAM model a rounding  $\mathbf{z} \in \{0, 1\}^n$  satisfying (3) can be computed, by the method of pessimistic estimators in  $O(Kn)$  time, such that the rounding error  $|(\mathbf{C}\mathbf{z})_k - (\mathbf{C}\mathbf{x})_k|$  is  $O(\max\{1, (\mathbf{C}\mathbf{x})_k\} \log K / \log \log K)$  for all  $k \in [K]$ .*

*Proof.* The theorem can be proved in the same way as in [10, Theorems 5 and 6].  $\square$

We now extend Theorem 4 to the case of  $\mathbf{C} \in ([0, 1] \cap \mathbb{Q})^{K \times n}$ . We make use of the derandomization presented in [4, 5], which consists in approximating  $\mathbf{C}$  by binary expansions and applying the method of pessimistic estimators to binary matrices (in our case applying Theorem 4).

**Theorem 5.** Let  $\mathbf{C} \in ([0, 1] \cap \mathbb{Q})^{K \times n}$  and let  $\mathbf{x} \in ([0, 1] \cap \mathbb{Q})^n$  be any solution that satisfies constraints (3). Then a solution  $\mathbf{z} \in \{0, 1\}^n$  can be deterministically computed in  $O(Kn \log n)$  time such that  $\mathbf{z}$  satisfies (3) and  $(\mathbf{Cz})_k$  is of value  $O(\max\{1, (\mathbf{Cx})_k\} \log K / \log \log K)$  for all  $k \in [K]$ .

*Proof.* The proof goes in the same manner as in [5, Theorem 2]. We present it here to show the idea and for the completeness. Set  $t = \lceil \log n \rceil$  and express each entry of  $\mathbf{C}$  in  $t$ -bit binary form. This can be done in  $O(Kn \log n)$  time. Let us denote by  $\tilde{\mathbf{C}}$  the  $t$ -bits representation of  $\mathbf{C}$ , namely  $\tilde{\mathbf{C}} = \sum_{l \in [t]} 2^{-l} \mathbf{C}^{(l)}$ , where  $\mathbf{C}^{(l)} \in \{0, 1\}^{K \times n}$ . Clearly,  $|c_{kj} - \tilde{c}_{kj}| < 2^{-t} \leq 1/n$ ,  $k \in [K], j \in [n]$ . Thus,  $\|\tilde{\mathbf{C}}\mathbf{x} - \mathbf{C}\mathbf{x}\|_\infty = \|(\tilde{\mathbf{C}} - \mathbf{C})\mathbf{x}\|_\infty \leq \|\tilde{\mathbf{C}} - \mathbf{C}\|_\infty \|\mathbf{x}\|_\infty \leq 1$  for every  $\mathbf{x} \in [0, 1]^n$ . We now apply Theorem 4 to the solution  $\mathbf{x}$  and the  $tK \times n$  binary matrix consisting of the rows of the binary matrices  $\mathbf{C}^{(l)}$ ,  $l \in [t]$  and get in  $O(tKn)$  time (and thus in  $O(Kn \log n)$  time) in the RAM model a solution  $\mathbf{z} \in \{0, 1\}^n$  satisfying (3) such that for all  $k \in [K], l \in [t]$  the rounding errors are as follows:

$$|(\mathbf{C}^{(l)}\mathbf{z})_k - (\mathbf{C}^{(l)}\mathbf{x})_k| \leq O(\max\{1, (\mathbf{C}^{(l)}\mathbf{x})_k\} \log(tK) / \log \log(tK)). \quad (6)$$

Therefore,

$$\begin{aligned} |(\tilde{\mathbf{C}}\mathbf{z})_k - (\tilde{\mathbf{C}}\mathbf{x})_k| &= \left| \sum_{l \in [t]} 2^{-l} ((\mathbf{C}^{(l)}\mathbf{z})_k - (\mathbf{C}^{(l)}\mathbf{x})_k) \right| \leq \sum_{l \in [t]} 2^{-l} |(\mathbf{C}^{(l)}\mathbf{z})_k - (\mathbf{C}^{(l)}\mathbf{x})_k| \\ &\stackrel{(6)}{\leq} O(\log(tK) / \log \log(tK)) \sum_{l \in [t]} 2^{-l} \max\{1, (\mathbf{C}^{(l)}\mathbf{x})_k\} \\ &\leq O(\log(tK) / \log \log(tK)) \sum_{l \in [t]} 2^{-l} (1 + (\mathbf{C}^{(l)}\mathbf{x})_k) \\ &\leq O(\log(tK) / \log \log(tK)) (1 + (\tilde{\mathbf{C}}\mathbf{x})_k). \end{aligned} \quad (7)$$

Let us analyze a performance of the solution  $\mathbf{z}$ :

$$\begin{aligned} (\mathbf{Cz})_k &\leq 1 + (\tilde{\mathbf{C}}\mathbf{z})_k \leq 1 + (\tilde{\mathbf{C}}\mathbf{x})_k + |(\tilde{\mathbf{C}}\mathbf{z})_k - (\tilde{\mathbf{C}}\mathbf{x})_k| \stackrel{(7)}{=} O(\max\{1, (\tilde{\mathbf{C}}\mathbf{x})_k\} \log(tK) / \log \log(tK)) \\ &= O(\max\{1, (\mathbf{Cx})_k\} \log(tK) / \log \log(tK)). \end{aligned}$$

The theorem follows when  $K \geq \log^{1/\gamma} n$  for some constant  $\gamma \geq 1$ , because it holds  $t \leq K^\gamma$ . We can satisfy this assumption by adding a number of zero cost (dummy) scenarios, if necessary. Alternatively, we can handle the case with a small number of scenarios (e.g.,  $K < \log^{1/\gamma} n$ , for  $\gamma \geq 2$ ) by applying some algebraic techniques (see, e.g., [5]).  $\square$

From Theorem 5 and the fact that  $\mathbf{x}^*$  is a feasible solution to  $\mathcal{LP}(L^*)$ , we immediately get the following corollary:

**Corollary 2.** MIN-MAX RS has a polynomial, deterministic  $O(\log K / \log \log K)$  approximation algorithm, where  $K$  and  $r_{\max}$  are parts of the input.

## 4 Concluding remarks

There is still an open question concerning the MIN-MAX REGRET RS problem, when both  $K$  and  $r_{\max}$  are parts of the input. For this problem, there exists a  $K$ -approximation algorithm,

known in the literature, and  $O(\log^{1-\epsilon} K)$  lower bound on the approximability of the problem, given in this paper. So, constructing a better approximation algorithm for MIN-MAX REGRET RS is an interesting subject of further research.

## Acknowledgements

This work was partially supported by the National Center for Science (Narodowe Centrum Nauki), grant 2013/09/B/ST6/01525.

## References

- [1] H. Aissi, C. Bazgan, and D. Vanderpooten. Min-max and min-max regret versions of combinatorial optimization problems: A survey. *European Journal of Operational Research*, 197:427–438, 2009.
- [2] S. Arora and C. Lund. Hardness of approximations. In D. Hochbaum, editor, *Approximation Algorithms for NP-Hard Problems*, chapter 10, pages 1–54. PWS, 1995.
- [3] V. G. Deineko and G. J. Woeginger. Complexity and in-approximability of a selection problem in robust optimization. *4OR - A Quarterly Journal of Operations Research*, 11:249–252, 2013.
- [4] B. Doerr. Generating Randomized Roundings with Cardinality Constraints and Derandomizations. In *STACS*, pages 571–583, 2006.
- [5] B. Doerr. Improved approximation algorithms for the Min-Max selecting Items problem. *Information Processing Letters*, 113:747–749, 2013.
- [6] A. Dolgui and S. Kovalev. Min-max and min-max (relative) regret approaches to representatives selection problem. *4OR - A Quarterly Journal of Operations Research*, 10:181–192, 2012.
- [7] A. Kasperski and P. Zieliński. On the approximability of minmax (regret) network optimization problems. *Information Processing Letters*, 109:262–266, 2009.
- [8] P. Kouvelis and G. Yu. *Robust Discrete Optimization and its applications*. Kluwer Academic Publishers, 1997.
- [9] M. Mastrolilli, N. Mutsanas, and O. Svensson. Single machine scheduling with scenarios. *Theoretical Computer Science*, 477:57–66, 2013.
- [10] P. Raghavan. Probabilistic Construction of Deterministic Algorithms: Approximating Packing Integer Programs. *Journal of Computer and System Sciences*, 37:130–143, 1988.
- [11] P. Raghavan and C. D. Thomson. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica*, 7:365–374, 1987.
- [12] A. Srinivasan. Approximation algorithms via randomized rounding: a survey. In M. Karoński and H. J. Prömel, editors, *Series in Advanced Topics in Mathematics*, pages 9–71. Polish Scientific Publishers PWN, 1999.